Integrated Water Quality and Aquatic Communities Protocol – Lakes and Ponds

Standard Operating Procedure (SOP) #13: Data Entry

Draft Version 1.0

Revision History Log:

Previous Version	Revision Date	Author	Changes Made	Reason for Change	New Version

Introduction

Data entry is a critical component of field studies that must be addressed to ensure the data collected are of good quality. This SOP provides details of how to complete data entry in the project databases, which include site characterization and fish. The database was developed using the NPS Natural Resource Database Template (NRDT) in Access 2007 and utilizes lookup tables and electronic forms that mimic the hardcopy datasheets to reduce transcription errors and allow for easy data entry. Data dictionaries for the database are provided in appendix M. Once data entry is completed, data are assessed for Quality Assurance/Quality Control processes, covered in SOP #16: Quality Assurance Project Plan. Once quality control processes have been implemented, data are uploaded into a master database that is used to conduct analysis and develop summaries for reports. The master database also has the ability to export the data into the NPSEDD (NPS Electronic Data Deliverables) format for integration into EPA STORET (STOrage and RETrival).

This data entry SOP describes a single process for one of two possible methods: (1) entering data into the project database, transcribing the data from paper field sheets; or (2) directly entering the data into a field Tablet PC during data collection. The latter is the preferred method; however, the methodology is identical. The same database is used since the Tablet PC operates a full version of Microsoft Windows operating system with the full functionality of a desktop computer.

Timelines and Responsibilities

Field crew members are responsible for completing the first two rounds of data validation and data entry. Data should be entered as quickly as possible following data collection. While it is the goal of the Network to collect data electronically wherever possible, data will occasionally be

recorded using datasheets. Hardcopy datasheets should be entered in the same week the data are collected. It is the responsibility of the Project Lead to ensure data are entered on a weekly basis and should schedule the field crew's time accordingly. The Project Lead should review the data after the first sampling period where field crews have been working on their own to ensure data are being documented properly. Periodically, the Project Lead should also review a subset of the data to ensure they are being recorded properly throughout the field season. In addition, prior to field crews starting a new park, the Project Lead should review the data to ensure they are accurate and complete.

Preparing the Database for Field Work

In order to prepare the database to be used in the field, the Project Lead will need to provide the Data Manager with a GIS layer of all the sites that will be visited during the upcoming season, a list of field crew members and contact information, and a list of all species expected to be encountered and identifiable in the field that will be recorded as part of this protocol (SOP #2: Field Crew Training). Note: the field species list need not include all potential species that might be subsequently identified in samples by contract laboratories. Once the Data Manager has the lists, he/she can begin to prepare the project database that will be used by the field crews that year.

Entering New Sites

To prepare the project database so it can be used in the field, there are a variety of steps (listed below) that need to be followed.

1. The first thing the Data Manager must do is load a list of sites that will be surveyed that year into the database.
    a. It is the responsibility of the Project Lead to work with the GIS Specialist to develop this list of sites.
    b. The final list of sites to be surveyed should be located at (to be done by the Project Lead): G:\Monitoring\Water_Quality_Monitoring\Lakes\Lakes_GIS\PARK\YYYY and the name of the file is PARK_YYYY.dbf. In both the file pathway and the file name, the "PARK" is the four letter park code and YYYY is the year of the survey.
    c. A copy of the site list file should be made and placed in: G:\Monitoring\Water_Quality_Monitoring\Lakes\Lakes_Data\Survey Areas\YYYY\PARK.
2. Open the front-end of the database that you have placed in your working directory.
3. Click on the [Administrative Tools] button.
4. Click the [Upload Sites] button.
5. Browse to the location of the file you created in step 1c above.
6. Click the [Upload Sites] button.
7. The sites table should now be populated with the following:
    a. Location_ID

b. Network
c. Park
d. Site_Name
e. Site_Type
f. GRTS Code
g. X-Coordinate
h. Y-Coordinate
i. Coordinate System
j. PLSS
k. Watershed
l. Subwatershed
m. County
n. USGS Map
o. GIS ID Number
p. GIS Shapefile Name

Entering Contact Information

Next, the Data Manager will need to enter the contact information for each of the individuals that could be working on this project, supplied by the Project Lead. To complete this task, follow the steps below.

1. Open the front-end of the database and click the [Lookup Tables] button.
2. Using the pick list, select tlu_contacts.
3. Delete any of the contacts that are not going to be involved in the project that year.

Add any new contacts (such as new crew members) that will be involved in the project by May 31st of that year. This information should be provided by the Project Lead as described in SOP #1: Preparations, Equipment, and Safety.

Updating the Pick List

Since this is a standardized protocol, the pick list values should not be changed, with the exception of a few rare occasions. It is the responsibility of the Project Lead to follow all change procedure processes associated with this protocol prior to having the Data Manager change one of these values. Once proper procedures have been followed, the Data Manager should follow these steps to update a pick list.

1. First, determine which fields need to be updated.
2. Click on the [Lookup tables] button on the main form of the database.
3. Using the pick list, select tlu_Enumerations.
4. Find the Enum_Group associated with the list of data you want to edit.
5. Delete any values you no longer want to use.
6. Add any new values that are needed. Be sure to complete all fields, including:
    a. Enum_Group exactly as it is in the database.
    b. Sort Order, which is the order you want the data to appear in the pick list.

   c. Enum_Code, the value that is stored in the database.
   d. Enum_Description, a description of the value that is stored in the database.
7. Once you are done, click the [Close] button in the upper, right corner of the form.
8. It will ask you if you want to save your changes, click [YES].
9. You are done.

Data Entry

It is the responsibility of the Project Lead to ensure that the laptop provided to the crew has the current version of the Klamath Network Mountain Lakes and Pond database.

To start, navigate to C:\Lakes\Lakes_Data. Within this folder, double click on the icon:  KLMN_FE_v1.00_20081124.mdb. This will open up a "gateway" to enter the data (Figure 1).

Procedures

1. On the left side of the database window that opens up is a tab named "Enter / edit data."
2. After clicking on "Enter / edit data," you will be prompted for setting some default values for this entry session (Figure 2).
3. Assure that the information is presented correctly. If you are a new user, follow the menu-driven process to add yourself. Note that name is last name_first name, with proper punctuation (capitalize first and last name, separated by an "underscore"). Set the correct park, datum, zone and protocol (NAD83, Zone 10N is default). Protocol will be Lakes Monitoring for all sites.
4. After clicking "OK," a list populated with existing data will appear (Figure 3). At the first data entry session, this list should be blank. You should be supplied with a blank database, the previous field season being stored and archived according to the Klamath Network Data Plan (Mohren 2007).
5. Click on "Add a new record" in the upper right hand corner.
6. Fill out all fields on the "Data Entry Form – New Record" screen that originates (Figure 4).
   a. First, in the location field, select the name of the lake that you are sampling. This should automatically populate the X/Y coordinates and the unit code fields.
   b. Next, enter the following fields:
      i. Start date – date you started the survey.
      ii. Arrival time – time you arrived at the lake.
      iii. Drive time – time taken from staging area/housing till parking for the start of hiking.
      iv. Hike time – time it took you to hike from your starting location (parking spot or camping site) to the lake.
      v. Descriptor of location – General (Forested, Alpine, sub-alpine).
      vi. Trails used to access – Provide a description of the trails you used to access the lake, beginning at the parking location or camp site.

SOP #13: Data Entry (continued).

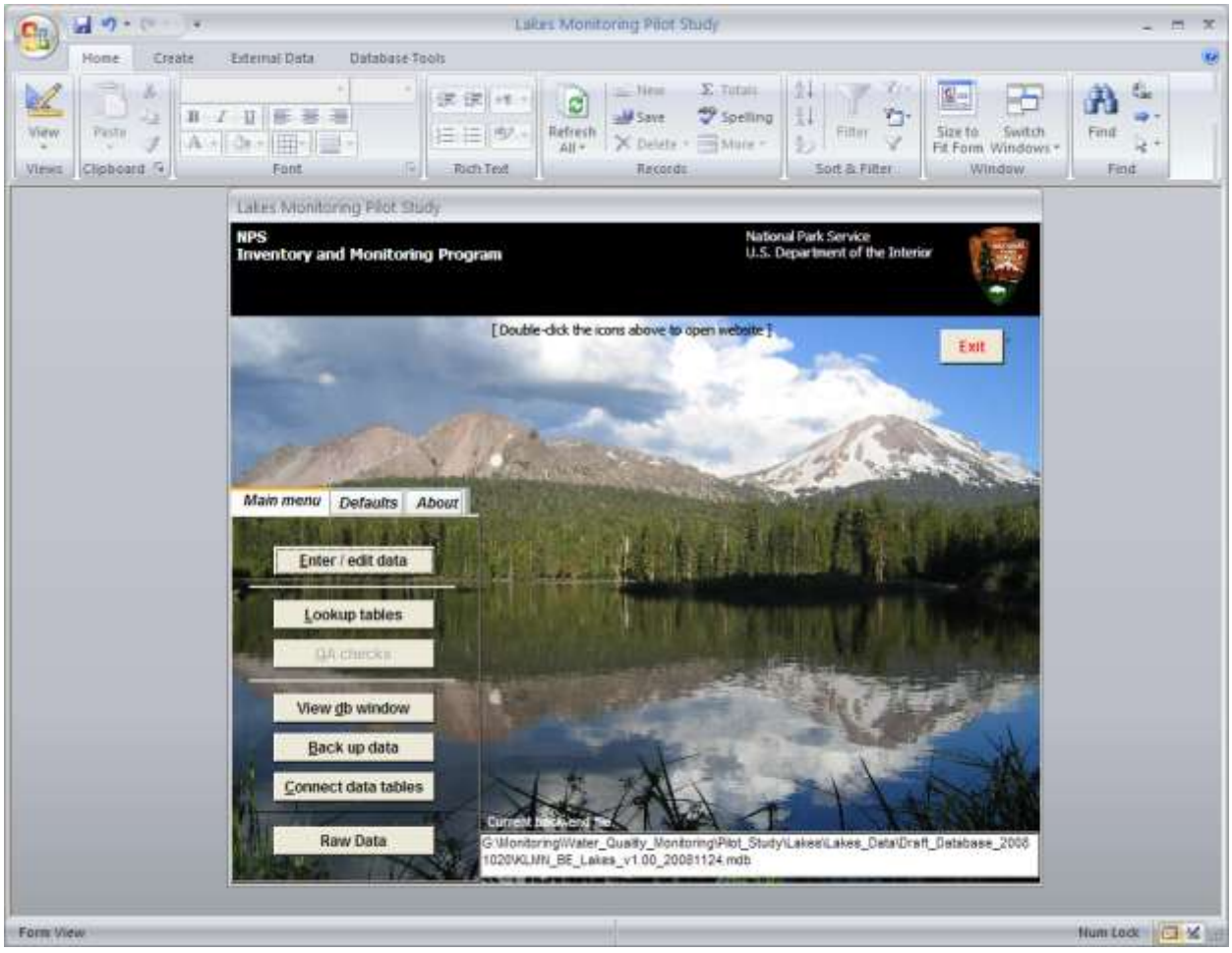


Figure 1. Klamath Network database gateway for entering Mountain Lakes and Ponds data.

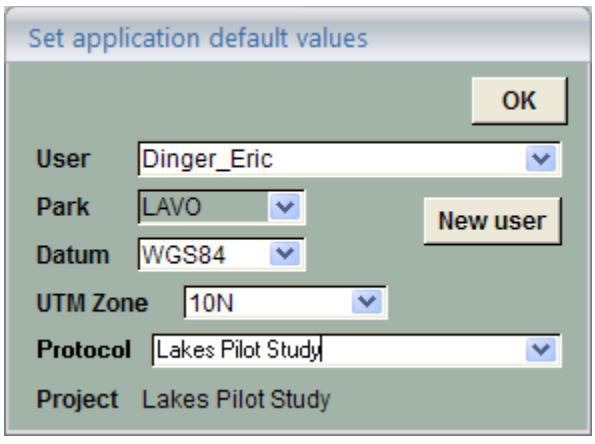
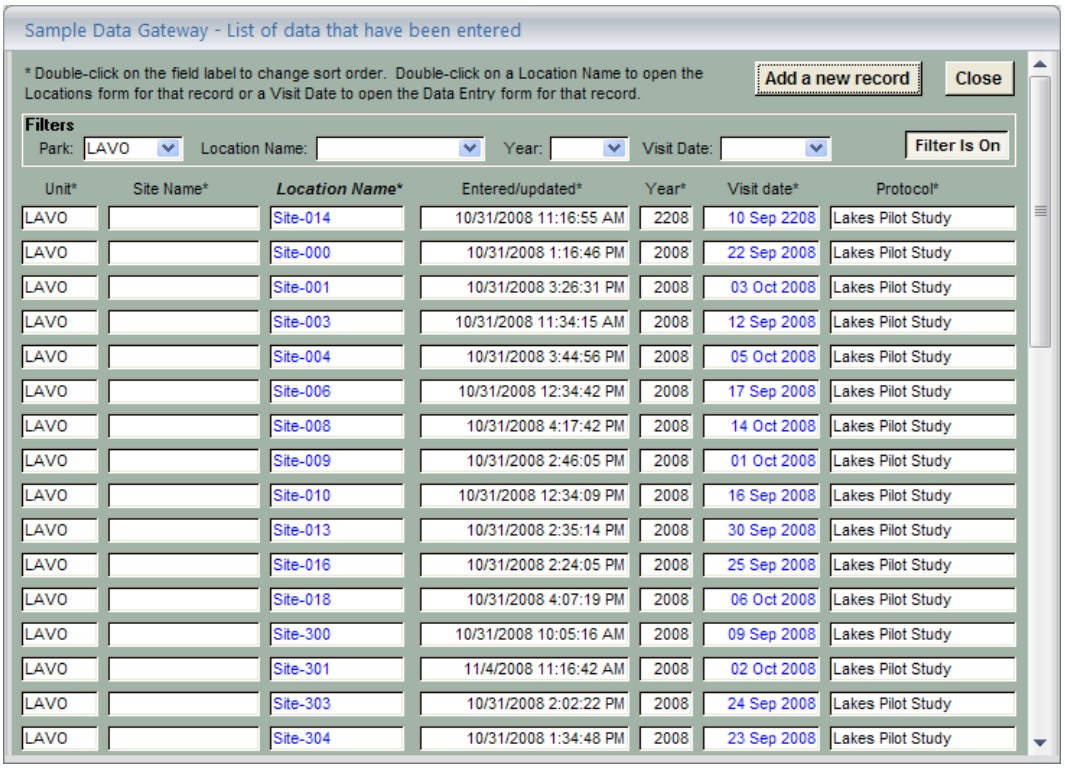


Figure 2. Default screen for beginning the data entry process.

SOP #13: Data Entry (continued).

7. Next, complete all the fields in the eight tabs that do not pertain to data collected on the Trimble unit or Manta data logger. Be sure to fill in all areas. Even if there appears to be a field that does not pertain to this particular site, be sure to note it with either N/A for not applicable or with a note in the "Visit notes" section. If necessary, "add a person" (e.g., a park specialist joined the crew for that day).
8. As data are being entered, the person entering the data should visually review each data form to make sure that the data on screen match the field forms.
9. Cycle through the tabs (highlighted in red in Figure 4) to ensure that all data are entered. At the conclusion of the data entry for this site, there should be **NO DATA** on the field form that is not entered into the database.
10. Because this database is designed to contain all data originating from this protocol, there will be fields not entered by the crew: Invertebrates, Water chemistry, Manta data, and Amphibians (embedded in the GIS shapefiles). **Note that there are some fill-able fields (e.g., the time and field split of invertebrates, and amphibian walk-around times). The crew must still enter these data!** These fields will be filled in by data import steps after the contract laboratories provide their data to the Project Lead. During data entry, tabs and fields to be imported later will be "hidden" from the database interface.



Unit*	Site Name*	*Location Name**	Entered/updated*	Year*	Visit date*	Protocol*
LAVO		Site-014	10/31/2008 11:16:55 AM	2208	10 Sep 2208	Lakes Pilot Study
LAVO		Site-000	10/31/2008 1:16:46 PM	2008	22 Sep 2008	Lakes Pilot Study
LAVO		Site-001	10/31/2008 3:26:31 PM	2008	03 Oct 2008	Lakes Pilot Study
LAVO		Site-003	10/31/2008 11:34:15 AM	2008	12 Sep 2008	Lakes Pilot Study
LAVO		Site-004	10/31/2008 3:44:56 PM	2008	05 Oct 2008	Lakes Pilot Study
LAVO		Site-006	10/31/2008 12:34:42 PM	2008	17 Sep 2008	Lakes Pilot Study
LAVO		Site-008	10/31/2008 4:17:42 PM	2008	14 Oct 2008	Lakes Pilot Study
LAVO		Site-009	10/31/2008 2:46:05 PM	2008	01 Oct 2008	Lakes Pilot Study
LAVO		Site-010	10/31/2008 12:34:09 PM	2008	16 Sep 2008	Lakes Pilot Study
LAVO		Site-013	10/31/2008 2:35:14 PM	2008	30 Sep 2008	Lakes Pilot Study
LAVO		Site-016	10/31/2008 2:24:05 PM	2008	25 Sep 2008	Lakes Pilot Study
LAVO		Site-018	10/31/2008 4:07:19 PM	2008	06 Oct 2008	Lakes Pilot Study
LAVO		Site-300	10/31/2008 10:05:16 AM	2008	09 Sep 2008	Lakes Pilot Study
LAVO		Site-301	11/4/2008 11:16:42 AM	2008	02 Oct 2008	Lakes Pilot Study
LAVO		Site-303	10/31/2008 2:02:22 PM	2008	24 Sep 2008	Lakes Pilot Study
LAVO		Site-304	10/31/2008 1:34:48 PM	2008	23 Sep 2008	Lakes Pilot Study

Figure 3. Date screen gateway for new record additions.

SOP #13: Data Entry (continued).

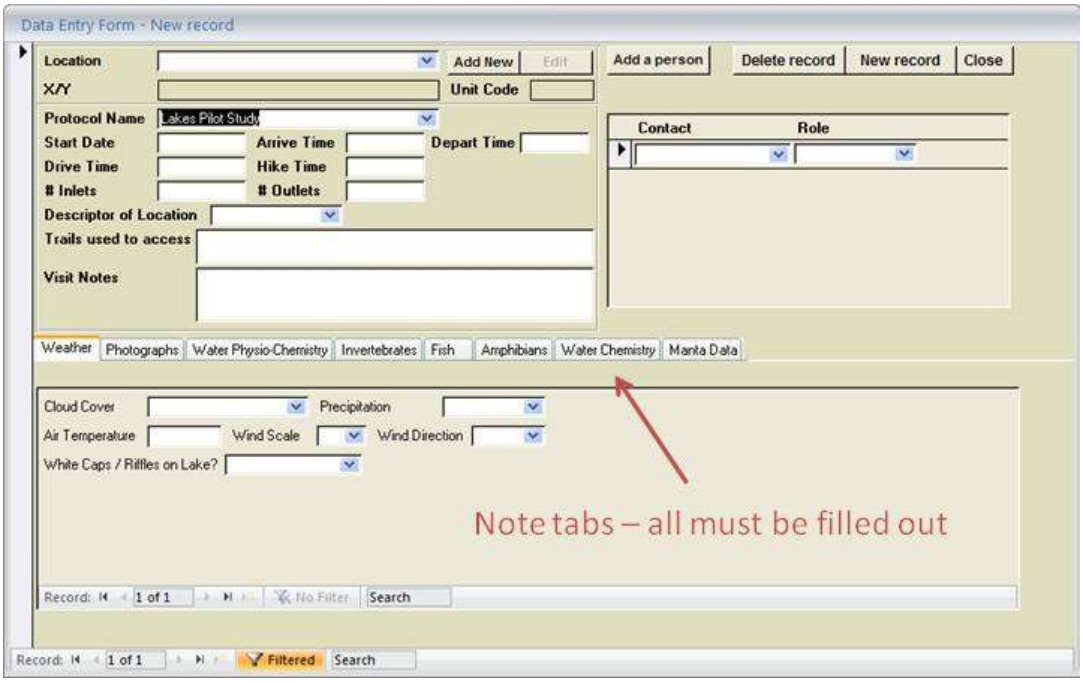


Figure 4. "Data Entry Form, new record" for entering Lake data. Note the tabs: the crew must cycle through each in turn to fill out all data.

11. Before concluding entering these site data, the person entering the data should **again** visually review each data form to make sure that the data on screen match those on the field forms.
12. If a new site ("add new") is necessary, the crew should fill out all applicable fields.
13. When done with the site, the person entering the data should click on "New record" to enter the next site.
14. At the conclusion of the data entry session, click on "Close" to exit the data entry screen.
15. "Exit" out of the main window; you are done.

Basic Guidelines

1. Enter all data exactly as they appear on the field form. If notes were taken short hand during the field visit, and it was your note, it is acceptable to write it as long-hand (i.e., grammatically correct). An example: On field form – "2 hikers seen." In database entry – "Two hikers were observed hiking by the site."
2. When entering data, no fields should be left blank. When a field is left blank by the observer that actually required data, enter -9 in that entry field. This value is used because -9 would not be an acceptable data value for any field. An example of when to enter -9 is if the crew ran out of alkalinity reagents and was unable to take a second (or first) reading. This should be noted until "visit notes."
3. When entering data, try to leave enough time to complete a block of data (e.g., the entire field visit). Do not partially enter a field visit.

4. After entering the data, write your initials and the date on the bottom of the field form in the applicable space.
5. If you have any questions about how to enter data, ask the Project Lead for clarification immediately; do not assume anything.

Data Verification and Validation

These are important steps from SOP #16: Quality Assurance Project Plan, but are repeated here to provide a comprehensive SOP for completing the data entry, including ensuring that the data are entered correctly.

Data Verification

Data verification of the data involves evaluating the correctness, conformance, compliance, and completeness of the entire dataset against the methods or procedures of the protocol (SWAMP 2008). Verification should be done on both field data (including field chemical analyses) and laboratory data (chemical, invertebrate, and zooplankton). It is the responsibility of both the Field Crew and Project Lead.

- Data Verification includes:
    - Visual Review at data entry – the technician verifies each value during input. Errors are corrected immediately.
    - Visual Review after data entry – after entry, data are printed out and compared to original hardcopy sheets.
    - Duplicate data entry – Randomly selected site data are entered as normal, but are duplicate records. Although time consuming in that it repeats data entry efforts, this gives an estimate of the data entry accuracy.
    - Review – It is the Project Lead's responsibility to review a subset of records to ensure that they are identical to the hardcopy datasheets.
    - For the duplicate data and review, the minimum number is 20% of the sites (approximately eight total sites).

Data Validation

After verification, the Project Lead reviews it against all criteria in the protocol, especially the QAPP criteria (e.g., holding times, laboratory duplicates, completeness goals, reporting limits). After successful validation, the Data Manager can send the data on to WRD for incorporation into NPStoret.

- Data Validation includes:
    - Data entry programming steps – The Project Lead, along with the Network Data Manager, will program steps design to prevent errors. For example, maximum depth of a lake entry will not allow the recording of a lake 100 m deep (since there are no lakes in Lassen Volcanic National Park that deep). This is an example

of a mistake that might occur if the technician accidentally enters "100" instead of "10.0."

- Outlier detection and review – Statistical review and graphical display will be used to detect outliers, which are unusually extreme values of a variable outside the range of normal values. In outlier review, it is important to realize that not all extreme values represent errors, but can reflect the real variation of the data in nature. Generally, outliers that cannot be ascribed to error will be flagged and retained.
- Review of "what makes sense" – The Field Crew Leader and Project Lead will compare and review the tabular data to confirm that everything "makes sense." Both should be intimately familiar with the types of data being collected, and as such should be able to detect mistakes. GIS data will be plotted and confirmed to match the spatial locations.

Database Backup

Data backups should be completed every day that new data are entered. Backups are created to save time in case of mistakes or database file corruption.

Backing up the database is a simple procedure.

1. From the main database gateway (Figure 1), click on "Back up data." You will then be prompted to confirm the backup. Alternatively, the user will be prompted to back up the database when exiting the program.
2. You will be prompted for a file location for the backup. Place it in: C:\Data\Lakes_Data\Backups. This location should be pre-created by the Project Lead or Data Manager on the laptop.
3. Save the file in this location with the default name given by MS Access. It will include a date and time stamp (e.g., 20090721_1133, for a backup made on the 21st of July, 2009 at 11:33AM).
4. Transfer the complete database and all backup files (including photos, multiprobe data, and GIS data [See SOP #12: Post-Site Tasks]) to the Project Lead, using a CD or flash drive. Deliver or mail to: Aquatic Ecologist (or Project Lead), Klamath Network, 1250 Siskiyou Blvd, Ashland, OR 97520, on a weekly basis.

Maintaining the Data Entry Log

A data entry log must be maintained during data entry (next page, also included in Appendix F). Each "data entry log" line item should be specific to a single data entry session. At the top of the data entry log, record the name of the data entry file and the names of the individuals entering data. During each data entry session, record the date, your initials, the field season year, the number of sites entered, the names of the sites, the number of hours spent entering data, and the name of the exported backup file). In addition, record any comments. Comments should include

the reason that "-9" was entered for data that did not conform to the database or unresolved edit needs.

Literature Cited

Mohren S. R. 2007. Data management plan, Klamath Inventory and Monitoring Network. Natural Resource Report NPS/KLMN/NRR—2007/012. National Park Service, Fort Collins, CO.

Surface Water Ambient Monitoring Program. 2008. Quality assurance project plan. California State Water Resources Control Board. Available online. (http://www.swrcb.ca.gov/water_issues/programs/swamp/docs/qapp/swamp_qapp_master090108a.pdf.) Accessed 3 December 2009.

SOP #13: Data Entry (continued).

Klamath Network Mountain Lakes and Ponds Data Entry Log

Data File:_____ Data Enterer Name:_____
Field Season Year_____ NOTE: Continue site names and comments on following lines as necessary.

Initials	Date	# of sites entered	Site names	# of hours to enter	Comments	Backup file name